Chapter 13
An "Eye" for an "I": The Challenges and Opportunities for Spotting Credibility in a Digital World

Jeff Hancock and Michael Woodworth

Rarely a day goes by without a new scandal that involves some kind of deception or fraud that has been perpetrated with the assistance of the Internet. Commonly, these deceptions involve people lying about who they are. These kinds of identity fraud stories are particularly compelling given the apparent ease with which individuals can craft a false identity or make false statements when they are hidden behind a computer screen. While lies about some aspect of one's identity or life are generally innocuous, this type of deception can also have disastrous consequences. The recent case of William Melchert-Dinkel illustrates this point. Melchert-Dinkel used a number of aliases, including posing as a suicidal female nurse named "li Dao" who actively encouraged individuals on the Internet to end their own lives. In a landmark decision, Judge Neuville referred to his actions as "lethal advocacy" and found that Melchert-Dinkel was guilty of aiding suicide in connection with the deaths of a Canadian female university student and an adult man in the UK.

Given that online environments consist of primarily text and, in some cases, photos, conventional wisdom considering credibility in the digital world dictates that it is easier to lie about who you are or what you are doing in comparison to face-to-face communication (see Colwell, Hiscock-Anisman, & Fede, present volume). Indeed, almost everyone has beliefs about how technology affects deception and deception detection. However, many of these beliefs are unwarranted and often the product of powerful psychological biases. But with billions of messages exchanged daily on the Internet regarding business, politics, national security, and interpersonal relationships, these misconceptions and errors can be deeply consequential.

We begin this chapter by examining some of the errors people make in judging deception and credibility online, including beliefs about how often people lie in

J. Hancock (✉)
Cornell University, New York, USA
e-mail: ith34@cornell.edu

M. Woodworth
University of British Columbia – Okanagan,
Vancouver, Canada

B.S. Cooper et al. (eds.), *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment*, DOI 10.1007/978-1-4614-5547-9_13,

digital contexts and the relative difficulty of detecting these deceptions. We then describe a number of studies we have conducted that empirically investigate people's ability to detect deception in text-based online communication, such as email, text messaging, online chat or status updates on Facebook, and what psychological factors appear to play a role. Given that the online world is rife with text, we describe a series of studies we have conducted looking at how the language of online messages can be mined to discover differences between truthful and deceptive messages. We also examine a similar but new line of work that examines how personality traits, such as psychopathy, can be potentially detected from language, and how this might play a role in how to judge credibility in online environments.

Finally, we end the chapter with a novel approach to credibility in digital contexts. The central idea revolves around the challenge of detecting deception after the fact. Given this difficulty, our approach suggests that structuring honesty may be a more fruitful approach for credibility online. In particular, we suggest using subtle primes and psychological constraints to reduce the likelihood that an individual will lie in a particular context. This approach aims to reduce the chance that deception will occur, rather than trying to detect it post conversation. We sketch out our initial thinking in this area and follow up with some promising new results.

Beliefs about Digital Deception

Most people believe that the Internet is awash with deception (Keyes, 2004), a belief that is supported by frequent media reports of people caught in a lie facilitated by some online technology. An apt observation is that, "on the Internet, no one knows you're a dog" (see Walther, 1996, p. 22). The idea that technology affords more deception than face-to-face interactions is not new to the Internet. Every time new technologies are developed that allow people to communicate at a distance, from the telephone and telegraph to the invention of the alphabet, the public has registered concerns about increases in lying (see Hancock, 2007). These beliefs about deception, however, run contrary to several studies that have examined how often people lie to each other online versus face to face or on the phone. In one series of diary studies that we conducted, participants reported lying least often in email (Hancock, 2004). In fact, in most studies, including DePaulo et al.'s (1996) seminal diary study on deception, the telephone is the medium of choice, with rates of lying in telephone conversations higher than typical Internet-based technology (see Hancock & Gonzales, in press).

The aforementioned results are inconsistent with the conventional wisdom that, when there are fewer cues, people should lie more often. We call this the *cue-availability heuristic* (Toma & Hancock, 2012) in which people assume that when there are fewer cues available in a communication medium, lies are more likely since fewer cues lower the chances of getting caught. According to this logic, people should lie more in text-based communication than in media that have more vocal or physical cues, such as the telephone or face-to-face. Why is it that people believe this to be the case when it comes to new technologies?

The most likely answer to this question relates to the recency of digital communication. Humans evolved communicating with one another when all members of the communication were physically present. This evolutionary trait has been in effect since humans began using speech, at least making it approximately 6–1,00,000 years old. Given this timescale, the ability to communicate at a distance is extremely new, with the invention of the alphabet around 3,000 years ago, and the use of email and text messaging only about 20 years old. Our systems, therefore, are potentially biased to be less trusting when not physically interacting with one another. Some of our own research described below calls into question how suspicious individuals are in online environments. This suggests that it may, in fact, be that individuals are still not comfortable deceiving in this relatively new communication medium—at least not at the rate that many people believe deception occurs online.

To examine the cue availability heuristic and its role in beliefs about deception and technology, we conducted a study that drew upon two well-established biases, one from psychology and one from communication studies. The first is the *double-standard effect*, in which individuals tend to believe that other people use deception more than they do themselves (Gordon & Miller, 2000). This perspective difference is due to biases in the way people perceive lies told by the self versus others, with lies told by the self perceived as more justified than those told by others. Saxe (1991) has also argued that this kind of self–other asymmetry in people’s beliefs about deception is critical to advancing our understanding of beliefs about deception, whether it be of the digital variety or not. The second bias is the *third person effect* (see Davison, 1983), which suggests that people do not like to perceive themselves as vulnerable to media effects because such an admission violates their sense of the self as in control of decisions and behaviors. For example, people tend to believe that advertising has a persuasive impact on others, but not on them. This effect has been demonstrated for many media dynamics (Perloff, 2002).

The question, then, is whether the combination of the third person effect and the double standard effect can explain people’s beliefs about deception in digital contexts. We (Toma, Jiang, & Hancock, under review) recently examined this question by asking participants in a national survey first about their beliefs about how often *other* people lie face to face, on the telephone, in email and text messaging (1 = not at all, 7 = all the time). We then asked them about their *own* lying behavior across these media. Finally, we asked them for their rationale, and whether cues were important and whether certain reasons, such as self-protection, played a role in their thinking.

The results revealed the double standard of deception—overall, people thought that others lied more than they did. But, the self-other difference was significantly larger when participants judged deception in email and text messaging, revealing a third person effect. That is, the double standard of deception is intensified for online media, with our participants believing that other people lie much more than they do in email and text messaging. When we asked them why they lie versus why others lie, the cue availability heuristic was apparent. People argued that other people use digital media to lie because there are no nonverbal cues to be detected and, therefore, they are less likely to be discovered than they are in face-to-face interactions. In contrast, participants argued that their own lying behavior was not driven by cues

but by justifiable reasons, such as protecting one's privacy. Thus, consistent with the third person effect, participants felt that others would be more affected by online media. Other people, relative to the self, were expected to engage in more deceptive behavior online, and for less noble reasons.

The aforementioned study makes clear that some people's beliefs about deception and credibility in online media are subject to important psychological biases, such as the self-other asymmetry that drives the double standard effect, and the media-based third person effect. Given these biases, it is important to look at how these kinds of biased beliefs may play a role in detecting deception in online contexts.

Detecting Deception Online

Considering the biases outlined above, are individuals *actually* more suspicious when interacting within computer-mediated contexts? They certainly appear to believe that, contrary to their own relatively low level of deception, others are engaging in a higher level of deception within online contexts. For example, Caspi and Gorsky (2006) found that 79% of participants believe that deception is widespread online while only 19% reported engaging in online deception themselves. In face-to-face situations, humans tend to err on the side of assuming that their conversational partner is telling the truth, and this truth bias appears to be difficult to extinguish (e.g., Vrij, 2008). It is believed that one of the numerous factors that may account for this truth bias is that many individuals erroneously believe that, if they can see a person, they will be able to detect deception (e.g., "I could tell by looking into their eyes"; Hancock, Woodworth, & Goorha, 2010; ten Brinke & Porter, present volume). Unfortunately, in many cases, individuals are not particularly good at detecting deceit, potentially inflating their confidence that others are being truthful (e.g., Porter, Woodworth, & Birt, 2000).

Woodworth, Hancock, Agar, Cormier, and Carpenter, (2010) examined the role of suspicion specifically in synchronous computer-mediated communications. One hundred and two undergraduate student dyads were asked to discuss four topics meant to approximate what they would typically discuss online (e.g., relationship issues and status and identity issues). One participant, the sender, was deceptive during two topics and truthful during the other two. Suspicion was also manipulated across three conditions ranging from low to high suspicion (e.g., "there is a strong likelihood that your communication partner is lying to you"). Deception detection accuracy was operationalized as the absolute difference between the sender's rating of their truthfulness (on a scale from 0, completely untruthful, to 10, completely truthful) and the receiver's rating of the sender's truthfulness (on the same scale). Surprisingly, level of suspicion did not significantly impact deception detection accuracy with all three levels of the suspicion manipulation achieving between 55 and 59% accuracy rates. Even when participants were led to be highly suspicious of their interaction partner, this did not positively impact their ability to detect deceit. Interestingly, the sample was comprised of psychology students who will typically

have more education and are likely to be more computer savvy than the general population. Independent of the suspicion manipulation, they should have presumably been more suspicious compared to individuals in the community.

One explanation parsimonious with face-to-face theory is that, regardless of the communication medium, it is far too stressful and incongruent with our instinct to always have to be guarded or wary of being deceived. It may be that this truth bias is ingrained at a level that has not trumped our common-sense knowledge of computer-mediated concerns (e.g., lack of nonverbal cues, or the fact that, in this particular experiment, individuals were interacting with anonymous strangers). Participants were far more likely to judge that the sender was truthful on all four of the topics (i.e., 61% of receivers) than deceptive on all four topics (i.e., only 1%). This demonstrated that lack of suspicion is concerning considering that, for an increasing number of individuals, computer-mediated communication is their primary means of social communication (e.g., Hancock, 2007). So, while individuals report generally believing that others are being deceptive more than themselves in online situations, they intriguingly appear to let their guard down during their own individual online interactions and defer to the truth bias, perhaps particularly in situations where they are familiar with the person they are interacting with and have built up some trust.

Unfortunately, because there is still an overall lack of comparison for how to accurately judge the veracity of information in online contexts, the default for being more sensitive for distressing information may be to assume that an individual is being deceptive about this “high stake” or serious information. The recent case of Cameron Moffat and Kruse Wellwood, who murdered 16 year-old Kimberly Proctor, is a perfect example of individuals being unable to correctly determine the honesty of the sender in an online context. Moffat admitted that, in the days leading up to the murder, he discussed different potential techniques for committing murder with “over a dozen people” in various online forums. One can only assume that, if he had similar conversations with these individuals in face-to-face contexts, somebody would have notified a parent or the authorities. However, their default assumption was likely that he was joking about such heinous information and would never have been honestly reporting intent to commit murder. Adding to the confusion, some of these conversations about the “real world” occurred within the online discussion group for the online fantasy game “World of Warcraft.” Sadly, if the police had been notified about some of these online interactions, this tragedy could have potentially been avoided. In addition to this surprising general lack of suspiciousness (and lack of clarity around the norms of what is honest intent within online contexts), there are a number of other issues that potentially compound the difficulty of detecting deceit in online contexts. A strong motivation to lie, such as in a high-stakes situation (e.g., punishment if caught), has always been considered one of the few variables that consistently serve to impede deceptive individuals (e.g., Depaulo, Kirkendol, Tang, & O’Brien, 1988). It is thought that the cognitive effort necessary to construct a lie when the individual is highly motivated may provide additional effective cues (considered to be largely nonverbal) for a receiver trying to ascertain the truth (see O’Sullivan, present volume). This has commonly been referred to as the *motivational impairment effect*. While the double standard effect discussed above suggests

that individuals often feel they have genuine and pure motivations to lie in an online environment, individuals are often highly motivated to tell a variety of harmful and nefarious lies as well.

Hancock et al. (2010) conducted the first empirical study to examine if the aforementioned well-documented impairment effect would translate to the computer-mediated context where dyads were asked to communicate with each other in instant messenger. Motivation was manipulated so that individuals in the "high" motivation category were informed that deception was a "very important skill" and that being able to successfully deceive was indicative of future successes in both employment and social contexts. The results indicated that, contrary to decades of face-to-face research, "high motivation" participants were the most successful at deceiving their partners, with increased motivation enhancing their success at deceiving. It would appear that a combination of the features available in the computer-mediated context is responsible for this novel finding. The exclusive availability of verbal cues (and lack of availability of nonverbal cues), combined with aspects of online interactions such as the opportunity to edit and plan out messages, may facilitate liars who are motivated enough to take advantage of these features. For example, any "late" or prolonged response latency in a face-to-face context has repeatedly been shown to be a perceived indicator of deception, potentially due to the extra time that is perceived that to be needed to craft a successful lie (e.g., Boltz, Dyer, & Miller, 2010), while in online communication, latencies of varying degrees are quite normal and expected. One has to wonder if the results would have been even stronger if the researchers had been able to manipulate motivation in a manner that more adequately approximated real-life situations. Presumably, many of the kinds of high-stakes lies that are told in online environments by sexual predators and other deceptive criminals would invoke a level of motivation that is difficult to create within an experimental paradigm.

Computer-mediated communication has become increasingly prevalent and, for many, is fast becoming the primary means of communication in some aspects of their life (Hancock, 2007). However, this type of communication is faced with a unique set of challenges for detecting deception, which include a general lack of suspiciousness and a set of features that benefit the goals of the highly motivated deceiver. Sadly, on the day she was murdered, Kimberly Proctor was lured and manipulated to meet up with Moffat and Wellwood by a number of text messages and no face-to-face communication. Fortunately, a number of studies outlined below suggest that, if the receiver is on the lookout for specific types of cues within the language of his or her communication partner, it may increase his or her confidence in the veracity of the information.

Linguistic Assessments of Deception Online

As outlined above, the digital world of communication is composed almost entirely of text. Almost all messages exchanged online involve a verbal message. Given the longstanding emphasis on nonverbal cues to assess deception, the textual nature of

the digital world seems to pose a challenge. Where there are challenges, however, there are also opportunities. In fact, one of the transformative aspects of digital communication for deception is that, unlike speech, everything that is communicated online leaves a digital trace. Even when speech is recorded, it still must be transcribed. Online, everything is already typed, producing a massive amount of text that can be analyzed.

Importantly, research on face-to-face deception has revealed important linguistic differences between deceptive and nondeceptive individuals. For example, Arciuli, Mallard, and Villar, (2010) found that individuals who were lying were significantly less likely to interject their speech with instances of “um” than truthful participants. The authors speculated that this type of speech utterance was associated with more natural and effortless speech, which would be difficult for many liars due to the cognitive stress of the lie. Interestingly, this potential linguistic clue to deception has also previously been found to increase during lying, presumably as an indicator of the increased cognitive demands of telling a lie (e.g., Vrij, Edward, Roberts, & Bull, 2000). The differences between face-to-face communication and computer-mediated communication (i.e., the parameters and affordances offered discreetly by each) offer an interesting opportunity to enhance our understanding of one by studying the other. For example, perhaps the nonverbal cues found in deceitful face-to-face communicators (e.g., Vrij, 2008) translate into textual evidence online. Further research on the topic would be welcomed as a way to better understand the similarities and differences between on and offline communication.

A substantial amount of recent work has also begun to examine the linguistic nature of deception in online contexts. This research has been driven by important advances in natural language processing, or the ability for computers to parse language, which is a potential boon to deception researchers. Researchers can now use computer programs to efficiently parse and count patterns in verbal messages, an approach that coincides with recent calls for researchers to focus more on verbal aspects of deception (Vrij, 2008). For example, a recent study by Duran, Hall, McCarthy, and McNamara, (2010) found that deceptive individuals were more likely to include redundancies, or the repetition of key information.

Consider one model that uses a very simple word-counting computerized approach to deception and language, the empirically derived Newman–Pennebaker (NP) model of deception. This model predicts several language features associated with deception, including fewer first person singular terms, fewer instances of exclusive conjunctions (e.g., words such as except, but, without) and more negative emotion terms (Newman, Pennebaker, Berry, & Richards, 2003). While this model was derived from controlled laboratory studies, this linguistic pattern has also been observed in deception by prison inmates (Bond & Lee, 2005) and, most recently, in courtroom testimonies of 46 defendants who were either found guilty of a crime and of perjury versus a group of defendants found guilty but who were later exonerated (e.g., in most cases by DNA evidence; Pennebaker, 2011). In this latter study, the strongest effects were from the use of first person singular pronouns. The more defendants used first person singular pronouns, the more likely they were to be innocent. This pattern suggests that use of first person singular

reflects ownership of a person's story. Use of exclusive words indicate that people are making a distinction between what they did do and what they did not do—essentially a marker of cognitive complexity.

We have also found similar patterns within laboratory studies (Hancock, Curry, Goorha, & Woodworth, 2008) and in political speech (Markowitz, Hancock, & Bazarova, 2011), and we have begun looking at how a variety of language processes, including negations, obligatory evidentiality, affect terms, coherence, and linguistic style matching markers can signal honesty/deception in text-based communication (see Hancock, 2004, 2007; Hancock and Gonzales, in press). In one project that shows how digital data can transform the analysis of deception beyond the Internet, we (Liu, Hancock, Zhang, Xu, Markowitz, & Bazarova, 2012; Markowitz et al., 2011) compared a corpus available on the Internet of false and non-false statements produced by officials in the Bush administration in the run up to the Iraq war. The false statements were identified by the non-partisan Center for Public Integrity, who used the 911 Commission conclusions that Iraq did not have weapons of mass destruction (WMD) or direct links to Al Qaeda at the time of the war, to identify a total of 535 false statements.

We applied the NP model of deception to the false and non-false statements collected by the Center for Public Integrity. Consistent with the model's predictions, false statements about WMD and links to Al Qaeda contained substantially and statistically significant reduced rates of first person singular ("I") and exclusive terms ("except, but") but contained more negative emotion terms and action verbs. Using this extremely simple model, we were able to classify approximately 76% of the statements correctly as either false or not false, suggesting that the language of the statements can predict whether or not the statement would turn out to be true of false. We have now begun examining other instances in which Western (i.e., English speaking) leaders made false claims and/or deployed misinformation (e.g., Churchill's deceptions during WWII).

It is tempting to begin to think of a set of consistently accurate verbal cues that predict deception, no matter what the context. For example, the decrease of first person singular across a wide range of studies suggests that it might be a reliable cue in verbal deception detection. While we believe that it is important to look at theoretically important cues regardless of the context, our research across a number of different studies has lead us to conclude that verbal cues are likely to be much more sensitive to contextual factors (such as the type of conversation, what the lie was about, whether the deception could be verified or is simply a person's opinion) than current assumptions around nonverbal behavior (see Ekman, 2001). We argue here that researchers should tailor their predictions for verbal cues to deception to the specific context, although we are still working to determine which key factors must be considered.

Consider, for example, three cues and how they operate across three very different studies. The cues are derived from the NP model (Newman et al., 2003): first person singular, which is expected to decrease during deception due to psychological distancing; conjunctives, which are also expected to decrease as deceptive language is often less complex than truthful; and more negative emotion terms, which should "leak" out given increases in anxiety around lying.

We have run three radically different studies to examine the aforementioned issues. The first was an experiment in which students chatted with each other over instant messaging about four topics, two truthfully and two deceptively (Hancock et al., 2008). In the second study, we looked at deceptive and truthful online dating profiles (Toma & Hancock, 2012). In this study, we examined how the free form text from the “about me” section changed with lies about the dater’s height, weight, and age. In the third study, we compared honest and deceptive hotel reviews (Ott, Choi, Cardie, & Hancock, 2011). Here, we asked one group of participants to write a five star review of a specific hotel as if they had actually stayed there and compared that to actual reviews of that hotel that presumably were honest.

What we found was that first person singular decreased, as predicted by the NP model, for both the chat and dating profiles deceptions, but actually increased for the deceptive hotel reviews. For conjunctions, we found that they decreased as expected for both the chats and the hotel reviews, but did not differ across deceptive and truthful dating profiles. Lastly, we found that negative emotion terms actually increased for both hotel and online dating deceptions but did not differ for the chats.

As we can see, the cues frequently differed across honest and deceptive accounts, but the differences may be systematic rather than random across the contexts. For instance, first person singular decreases when people may feel guilty about their deception, which might be the case in the deceptive chats and dating profiles where our participants were lying about aspects important to the self, such as important beliefs and identity. In contrast, in the hotel reviews, the whole point of the lie was to convince readers that they were actually there; thus, the liars over-emphasized first person singular. Conjunctives appear to be sensitive to how cognitively demanding the deception is. This is likely the case for the hotel reviews, which would require recreating a scene and an experience and, for the chats, which would require lies in real -time. In contrast, when creating an online dating profile and lying about aspects of the self, in which one has the time to construct and edit the well-known topic of the self, the cognitive demand should be moderated. Lastly, negative emotion terms appear to be sensitive not only to “leaked” emotion, but also to be a strategically deployed cue. Negative emotion terms were reduced in the two contexts in which the lies involved “selling” something (over and above the deceit in and of itself) – either how attractive the dater was or in how wonderful a hotel is. While it is impossible to know if these post hoc speculations can explain the pattern of results across these three different studies, we argue in this chapter that it is critical to consider how psychological dynamics and objectives differ across deceptions.

Taken together, we believe that the ability to analyze texts from a variety of domains points to a context-dependent approach to deception online. Deception researchers should consider the context when developing predictions about verbal cues, rather than trying to identify universal cues of deception that should apply to every context (see O’Sullivan, present volume). Why, for example, would we expect deceptions in an insurance fraud to be the same as deceptions about hotel reviews or about rationales for taking a country to war? In the digital world and with new tools

for parsing language, we should instead consider the specific circumstances and intent when constructing theoretically derived verbal predictions.

The Potential Role of Personality for Online Deception and Manipulation

When considering other context-driven variables that may impact credibility in online environments, the personality of both the deceiver and the individual being deceived are also important to keep in mind. An individual's language is arguably one of the best ways to glean important insights into his or her thoughts and beliefs. An increasing number of research projects have utilized automatic linguistic analysis programs to examine the language of other types of clinical populations and found that they can successfully differentiate between a variety of individual factors (e.g., Tausczik & Pennebaker, 2010). Previous research suggests that language may reveal important insights into both the personality and psychological make-up of an individual. Oberlander and Gill (2006) conducted an automated analysis of the email communication of a group of students and found a number of consistent linguistic style patterns based on the personality of the participant. For instance, a higher level of extraversion was associated with a preference for adjectives, whereas lower levels of neuroticism were linked to a preference for adverbs (see also Pennebaker, Mehl, & Niederhoffer, 2003).

Until recently, no automated language analysis programs had been employed to analyze the speech production of criminals and, more specifically, of psychopathic offenders. Previous studies that employed human coders have suggested particular language characteristics of psychopathic offenders. For example, Porter and Woodworth (2007) found that individuals scoring higher on psychopathy were more likely to exaggerate the reactivity of the homicide they committed and to omit some core detail of the incident than those scoring low on psychopathy. However, using automated language programs is arguably preferable in some cases, considering that many of the aspects of language measured with these programs are not consciously controllable by the speaker or measurable by human coders. Further, they are arguably more efficient than human coders both in terms of consistency and speed by which large amounts of text can be analyzed. Psychopaths are known to be particularly skilled at manipulating, deceiving, and controlling their self-presentation, making an automated enquiry into their language production another way to potentially obtain important insights into their behavior. Further, if they are demonstrating particular types of language patterns, it might be possible to more readily detect them in online environments where the vast majority of information will be text based.

Hancock, Woodworth, and Porter, (2011) used text analysis tools to examine the crime narratives of 14 psychopathic and 38 non-psychopathic homicide offenders. Psychopaths showed reliable differences relative to their nonpsychopathic counterparts such as focusing more on material needs during their narratives (e.g., food,

drink, money) and making fewer references to social needs (e.g., family, religion/spirituality). Psychopaths also used more past tense and less present tense verbs in their narratives, suggesting a greater psychological and emotional detachment from the incident. Consistent with the above, their language was less emotionally intense and pleasant.

The above study was one of the first to suggest that language may be used as a red flag by certain types of aversive personalities; in this case, the psychopathic personality, who is known to have a penchant for manipulation lying, and an ability to sense weaknesses (such as fear) in other individuals (e.g., Woodworth & Waschbusch, 2008). Interestingly, Wheeler, Book, and Costello, (2009) found that individuals who possess a high number of psychopathic traits were also better able to discern more vulnerable individuals from less vulnerable individuals, based on gait and other nonverbal cues. Further, individuals with a particularly concerning combination of personality characteristics known as the Dark Triad (Paulhus & Williams, 2002) which is a combination of subclinical psychopathy, narcissism, and Machiavellianism engage in the manipulation of others and the use of exploitation (Jonason, Li, & Teicher, 2010). Black, Woodworth, and Porter, (in preparation) are conducting one of the first research projects that explores whether Dark Triad individuals also will have an enhanced ability to detect vulnerability in individuals, as well as the verbal and nonverbal cues that they use to detect vulnerability. Once researchers possess a deeper understanding of the cues that Dark Triad individuals use to detect vulnerability in face-to-face interactions, this knowledge will lead to an investigation to determine whether exploitative and deceptive individuals are able to detect vulnerability in an online setting without the presence of any traditional nonverbal cues.

These types of studies lead to the troubling question of whether certain personalities or individuals are actually more prone to being preyed upon or deceived in online environments. For example, face-to-face research has demonstrated that some individuals are more vulnerable to being taken advantage of than others due to their own personality traits, such as low self-esteem and low assertiveness (e.g., Egan & Perry, 1998). Whether or not similar results would be obtained in an online environment remains to be seen. However, it is important for individuals interacting within computer-mediated domains to both create an environment and present in a manner where they can be most confident of the veracity of the sender of information.

Structuring Honesty—Promoting Truth Versus Detecting Deception

As has been made clear by numerous meta-analyses (e.g., Bond & DePaulo, 2006), deception detection is difficult for humans, who often perform effectively at chance in laboratory settings. Although much of the literature has focused on assessing credibility or detecting deception, another approach that might be useful is reducing

the likelihood that an individual will *produce* a lie when given the opportunity. Given that deception detection is difficult, researchers should focus on reducing the chance of a lie before it occurs.

How might the aforementioned be accomplished? One approach is to prime honest behavior. Evidence from evolutionary psychology suggests that pro- and anti-social behaviors can be manipulated using subtle primes. For example, in one study, researchers alternated a photo placed on a cup used to collect donations for the use of cream for coffees (Bateson, Nettle, & Roberts, 2006). One week, the photo had a pair of human eyes; the next, the photo was of flowers. At the end 10 weeks, the eye photo cup had collected significantly more than the flower cup. In another study, Haley and Fessler (2005) found that simple cartoons of eyes could prime a sense of surveillance and enhance cooperation in a dictator game. In particular, when players had two black dots over one dot on their computer screen, which represents two eyes and a nose, they gave more money to their partner in a money-splitting game than when there was one dot over two dots. In conversations, social psychologists have been able to prime more polite and more rude behavior by manipulating the kinds of words used in a conversation (Chartrand & Bargh, 1999).

These studies, while not focusing on deception per se, but the larger category of dishonest behaviors, suggest that individuals might be primed to be more honest in a certain situation where honesty is particularly important, such as in a witness report or a resume. Thus, we argue that the digital environment could be modified to prime more honest behavior. Imagine a witness report for an insurance claim that is filled out online. On the form could be placed the two dots above one dot configuration, perhaps as a logo, which has been shown to prime more pro-social behavior. Could it also prime the witness to be more honest in completing their report? If this was the case, the applications seem endless given the wide range of human activities now conducted online.

A second approach to enhancing honesty and credibility online would be to attempt to constrain people's ability to lie. One important lack of constraint in some digital contexts is that people can behave anonymously, such as in Internet chat rooms. But, in many other digital domains, there are connections between the person's virtual behavior or identity and their real-world identity—these connections are called *warrants* (Walther & Parks, 2002). Facebook, for example, made it clear to users from the beginning that their profiles should be for real individuals, and they initially implemented this policy by requiring an email from a university domain (e.g., @harvard.edu). Because Facebook profiles are tightly connected to their real-world identities, they should be credible and accurate. Recent research suggests this is the case. Back et al. (2010) found that individuals can accurately assess other individuals' personality traits using only Facebook information about that individual. Other work has found that the more warranted an identity is in an online space (e.g., photo, real name, presence of real-world friends), such as email or social networking sites, the more honest that person reported being in that space (Warkentin, Woodworth, Hancock, & Cormier, 2010).

Taken together, these studies suggest that the communication environment online can be manipulated to increase the degree to which people produce credible, honest behavior online. First, primes can plausibly be inserted into an environment that

should lead to more honest behavior. Second, warrants that connect an individual to their real-world identity should lead to more honest behaviors and credible communication than unwarranted situations. However, certain personality types, such as those scoring high on psychopathy (or on the dark triad), will likely be much more resistant to conventional means that may attempt to appeal to their conscience or empathy as a human being.

Conclusion

In summary, a review of deception in computer-mediated communication reveals that there are important implications across a variety of online communication settings. Many individuals are now conducting a substantial amount of their social interactions online, and often appear to be willing to divulge an inordinate amount of personal information. This is particularly true for teenagers, and even children who are still in the 10–12-year-old age range (e.g., Lenhart, Purcell, Smith, & Zickuhr, 2010). Business interactions and networks have also become increasingly geared toward online communication (e.g., Logsdon & Patterson, 2009). Identifying deception still poses unique challenges in online environments. For example, there is a lack of social norms available for what even constitutes deception. The recent divorce proceedings of Amy Pollard and her spouse David help to illustrate just how difficult this task may be. Pollard accused her spouse of engaging in what she believed was serious deceptive behavior online. She caught her husband engaging in online sexual activity between his avatar and another participant’s female avatar (i.e., a virtual call girl), and believed this to be tantamount to cheating. Based on the fact that her spouse had engaged in digital adultery, she filed for divorce citing “unreasonable behavior.”

Everyone will have a different opinion regarding both the seriousness of this behavior as well as whether this would constitute deceptive behavior truly indicative of infidelity. Interestingly, although it is clear that Pollard felt she had been deceived, in this case, what is unclear is the type (or nature) of deception that her (ex) husband had actually engaged in. Understanding the veracity (or seriousness) of the information provided in computer-mediated contexts was also a frustrating challenge in the Kimberly Proctor murder case outlined above. Further, it appears that the truth bias that is so evident in face-to-face environments is also present in online environments, despite individuals’ expectations that others will lie more often (and for less selfless reasons) than themselves. Further complicating the matter, highly motivated individuals, who have unparalleled access to potential victim pools in online environments, appear to benefit from features inherent in online communication, as well as the lack of traditional nonverbal cues.

Despite these troublesome aspects of online communication, the increasing prevalence of computer-mediated communication also affords many chances for us to improve our understanding of both deception as well as social interaction. Research conducted in online contexts discussed in this chapter has demonstrated that the

type of language deceptive individuals will produce will vary across both context as well as the motivations of the deceiver. Further, it would appear that certain personality types, such as psychopaths, engage in specific patterns of language use that may facilitate their detection both in online and face-to-face environments. Research is also beginning to suggest particularly effective parameters that could be employed to facilitate honest communication online. These include creating an environment that requires individuals interacting online to provide a variety of warrants to decrease their feelings of anonymity. Priming individuals in computer-mediated communication with social cues that in face-to-face contexts have been effective in instilling increased responsibility (e.g., including a simple image of being watched) may also be effective for reducing the amount of deceptive behavior online. While deception may currently be posing unique challenges for online communication, this environment also arguably provides us with distinct opportunities to improve our ability to understand both the mechanics of deception as well as parameters aimed at increasing our success at accurately gauging deceit.

References

Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth telling in speech. *Applied Psycholinguistics, 31*, 397–411.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S.D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*, 372–374.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*, 412–414.

Black, P. J., Woodworth, M., & Porter, S. (in preparation). The influence of the dark triad on the ability to detect vulnerability in others.

Boltz, M. G., Dyer, R. L., & Miller, A. R. (2010). Are you lying to me? Temporal cues for deception. *Journal of Language and Social Psychology, 29*, 458–466.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214–234.

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology, 19*, 313–329.

Caspi, A., & Gorsky, P. (2006). Online deception: Prevalence, motivation, and emotion. *Cyberpsychology & Behavior, 9*, 54–59.

Chartrand, T. L., & Bargh, J. A. (1999). The chamelion effect: The perception–behavior link social interaction. *Journal of Personality and Social Psychology, 76*, 893–910.

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly, 47*, 1–15.

DePaulo, B. M., Kirkendol, S. E., Kashy, D. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replication and extension. *Journal of Nonverbal Behavior, 12*, 177–202.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied PsychoLinguistics, 31*, 439–462.

Egan, S. K., & Perry, D. G. (1998). Does low self-regard invite victimization? *Developmental Psychology, 34*, 299–309.

Ekman, P. (2001). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York, NY: Norton & Company, Inc.

Gordon, A. K., & Miller, A. G. (2000). Perspective differences in the construal of lies: Is deception in the eye of the beholder? *Personality and Social Psychology Bulletin, 26*, 46–55.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody’s watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*, 245–256.

Hancock, J. T. (2004). Verbal irony use in computer-mediated and face-to-face conversations. *Journal of Language and Social Psychology, 23*, 447–463.

Hancock, J. T. (2007). Digital deception: When, where and how people lie online. In K. McKenna, T. Postmes, U. Reips, & A. N. Joinson (Eds.), *Oxford handbook of internet psychology* (pp. 287–301). Oxford: Oxford University Press.

Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. T. (2008). On lying and being lied to: A linguistic analysis of deception. *Discourse Processes, 45*, 1–23.

Hancock, J.T. & Gonzales, A. (in press) To lie or not to lie online: The pragmatics of deception in computer-mediated communication. In S. Herring, D. Stein, & T. Virtanen (Eds.) *Handbook of pragmatics of computer-mediated communication*. Berlin, Germany: Mouton de Gruyter.

Hancock, J. T., Woodworth, M. T., & Goorha, S. (2010). See no evil: The effect of communication medium and motivation on deception detection. *Group Decision and Negotiation, 19*, 327–343.

Hancock, J. T., Woodworth, M., & Porter, S. (2011). Hungry like the wolf: A word pattern analysis of the language of psychopaths. *Legal and Criminological Psychology*. doi:10.1111/j.2044-8333.2011.02025.x.

Jonason, P. K., Li, N. P., & Teicher, E. A. (2010). Who is James Bond? The dark triad as an agentic social style. *Individual Differences Research, 8*, 111–120.

Keyes, R. (2004). *The post-truth era: Dishonesty and deception in contemporary life*. New York, NY: St. Martin’s Press.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social media & mobile internet use among teens and young adults. In *Pew Internet & American Life Project*. Retrieved from http://www.pewinternet.org/Reports/2010/Social-Media-and-Young-Adult.

Liu, X., Hancock, J. T., Zhang, G., Xu, R., Markowitz, D., & Bazarova, N. (2012). *Exploring linguistic features for deception detection in unstructured text*. Presentation at the Proceedings of the International Conference on System Sciences, Hawaii, USA

Logsdon, J. N., & Patterson, K. D. W. (2009). Deception in business networks: Is it easier to lie online? *Journal of Business Ethics, 90*, 537–549.

Markowitz, D., Hancock, J. T., & Bazarova, N. (2011). *The language of presidential lies: How words can reflect lies about war, personal scandal and state secrets*. Presentation at the 97th Annual Meeting of the National Communication Association, New Orleans, LA.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665–675.

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Process, 42*, 239–270.

Ott, M., Cardie, C., Choi, Y., & Hancock, J.T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011)*, 309–319.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality, 36*, 556–563.

Pennebaker, J. W. (2011). *The secret life of pronouns*. New York, NY: Bloomsbury Press.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*, 547–577.

Perloff, R. (2002). The third person effect. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 489–505). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Porter, S., & Woodworth, M. (2007). “I’m sorry I did it … but he started it”: A comparison of the official and self-reported homicide descriptions of psychopaths and non-psychopaths. *Law and Human Behavior, 31*, 91–107.

Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior, 24*, 643–658.

Saxe, L. (1991). Lying: Thoughts of an applied social psychologist. *American Psychologist, 46*, 409–415.

Tausczik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24–54.

Toma, C., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication, 62*, 78–97.

Toma, C., Jiang, C., & Hancock, J. T. (under review). The deception-media double standard: Self-other asymmetry in beliefs about deception across media. *Cyberpsychology, Behavior and Social Networking*.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. West Sussex, England: Wiley.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior, 24*, 239–263.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–43.

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). Thousand Oaks, CA: Sage.

Warkentin, D., Woodworth, M., Hancock, J.T., & Cormier, N. (2010). Warrants and deception in computer mediated communication. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW2012)*, 9-12.

Wheeler, S., Book, A., & Costello, K. (2009). Psychopathic traits and perceptions of victim vulnerability. *Criminal Justice and Behavior, 36*, 635–648.

Woodworth, M., Hancock, J., Agar, A., Cormier, N., & Carpenter, T. (2010). *Suspicion in synchronous computer-mediated communication: Preliminary results*. Presentation at the Proceedings of the International Conference on System Science, Hawaii, USA.

Woodworth, M., & Waschbusch, D. (2008). Emotional processing in children with conduct problems and callous/unemotional traits. *Child: Care, Health and Development, 34*, 234–244.